LIFE Project Acronym and Number

ANTARES

LIFE08 ENV/IT/000435

Deliverable Report

Deliverable Name and Number

Deliverable 16

Evaluation on the model performances on at least 20 models (preliminary report)

Deliverable Date

31/08/2011

Deliverable Data

Associated action	Action 4: list of (Q)SAR models for the ecotoxicological, toxicological and environmental endpoints for REACH, and their rewiev
Beneficiary Responsible	KnowledgeMiner Software Frank Lemke
Monitoring	IRFMN

The scope and procedure of Deliverable 16

This deliverable takes the output of Deliverable 13, which, within the same Action 4, listed the QSAR models available for all the endpoints mentioned by REACH. For some endpoints no model was available, but within Deliverable 16 we focus on the best models available.

From Deliverable 13 a series of candidate endpoints can be extracted, because for these models there were models referring to a quite large number of compounds, and because a reasonable number of models exist.

In order to provide a wide basis of the use of the QSAR models, four different categories of endpoints have been used, following the scheme of endpoints for REACH: physico-chemical, environmental (fate, etc), ecotoxicological, and toxicological endpoints. At least one endpoint for each category should be ideally chosen, 4 in total, and, in order to give preference to endpoints which use more animals for the experiments, we chosen 4 other enpoints, two of them referring to human toxicology. Thus, the overall number of endpoints is 8, for this preliminary report.

With the target to get 8 endpoints, these are the first choise resulting from Deliverable 13:

- 7.8 Partition Coefficient N-Octanol/Water (phys-chem)
- 8.9.1 Carcinogenicity Study (human tox)
- 8.4.1 In-Vitro Gene Mutation Study In Bacteria (human tox)
- 8.5.1. Acute Toxicity-By Oral Route (human tox)
- 9.1.1 Short-Term Toxicity Testing On Invertebrates /Daphnia (ecotox)
- 9.1.3. Short-Term Toxicity Testing On Fish (ecotox)
- 9.2.1.1. Ready Biodegradability (environ. behaviour)
- 9.3.2. Bioconcentration Factor (environ. behaviour)

In this list we also indicated if the endpoint is a physico-chemical, toxicological,ecotoxicological, environmental behaviour endpoint. The numbering before the endpoint refers to the REACH regulation.

We mention that even if a certain endpoint does not use animals, nevertheless it may provide useful information, and in some cases it may even represent a surrogate of an endpoint using animals. For instance, this is the case of logP, the endpoint 7.8. LogP can be used to waive animal testing for the definition of the bioconcentration potential of a chemical substance, according to REACH.

Within this deliverable we analysed the most promising QSAR models for these 8 endpoints, and evaluated if they can be proposed for further screening within the following actions, aimed to verify their possible use within REACH.

In order to select these most promising models, we used the criteria identified within Action 2, Deliverable 2. First, we used the main crieria identified within Deliverable 2, and in some cases,

since there was a quite large number of models with similar total scores, we used the additional criteria.

These criteria have been used to rank the models for the same endpoint, and thus to identify the models which more likely can be used for REACH. However, this evaluation, is preliminary and is only the starting phase of the evaluation. Indeed, this evaluation is done only on a "nominal" basis, considering some general criteria, within this deliverable mainly as declared by the model developers. Thus, the scores of the tables used within this deliverable should NOT be used as a quality evaluation of the models. Further check will be done in the successive Actions. The target is to get at least 25 QSAR models to be proposed for Action 5. Preferably, from this deliverable a larger number of models should be obtained, in order to have more possibilities to find the best models for further check.

We notice that the implemented software offers advantages compared to the models which are only described in the scientific literature, because the descriptors, equations, and model outputs are standardised and reproducible; indeed they have been implemented into a tool. In case of models from the scientific literature the scores relative to these parameters are typically low. Thus, most of the preferable models arise from those available through the internet, or commercial sources.

Below we will analyse the QSAR models and discuss them.

7.8 PARTITION COEFFICIENT n-Octanol/Water (logP)

Software (and developer):

Freely available:

- ✓ EPISUITE (U.S. EPA) http://www.epa.gov/opptintr/exposure/pubs/episuite.htm;
- ✓ SPARC (U.S. EPA) http://archemcalc.com/sparc;
- ✓ VCCLAB (Virtual Computational Chemistry Lab) http://www.vcclab.org/.

Commercial:

- ✓ ACD/ADME Suite with AbSolv module (ACD Labs) http://www.acdlabs.com/;
- ✓ ADMETox/Pallas (CompuDrug) http://www.compudrug.com/;
- ✓ ADMET Predictor (Simulations Plus Inc.) http://www.simulations-plus.com/;
- ✓ ASTER (U.S. EPA) http://www.epa.gov/med/Prods_Pubs/aster.htm/;
- ✓ ADMEWORKS including Predictor and ModelBuilder (Fujitsu) http://www.fqs.pl/;
- ✓ ChemOffice (CambridgeSoft) http://www.cambridgesoft.com/;
- ✓ ChemProp (Helmholtz Centre for Environmental Research, UFZ) http://www.ufz.de/;
- ✓ ClogP (DAYLIGHT) http://www.daylight.com/;
- ✓ ChemDBsoft with MOLPRO Package including SLIPPER (ChemDBsoft) http://www.chemdbsoft.com/;
- ✓ ChemSilico Predictors, i.e. CS LogWS/D/P, CS BBB/PB/HIA (ChemSilico) http://chemsilico.com/;
- ✓ Jchem with Calculator Plugins (ChemAxon) http://www.chemaxon.com/;
- ✓ KnowItAll ADME/Tox (Bio-Rad Laboratories) http://www.bio-rad.com/;
- ✓ MetaDrugTM (Genego) http://www.genego.com/;
- ✓ MolCode Toolbox http://molcode.com/;
- ✓ Molecular Modeling Pro (ChemSW) http://www.chemsw.com/molecularmodeling.htm/;
- ✓ NorayMet ADME (Noray Bioinformatics) http://www.noraybio.com/;
- ✓ Pipeline Pilot (Accelrys Scitegic) http://accelrys.com/;
- ✓ PreADME (Bioinformatics and Molecular Design Research Centre) PreADMET web-based application (BMDRC) http://www.bmdrc.org/;
- ✓ ProPred (CAPEC) http://www.capec.kt.dtu.dk/;
- ✓ TerraQSAR™ - http://www.terrabase-inc.com/.

Overview of the evaluation of the QSAR models for logP

logP	EPISuite	VCCLAB	SPARC	ADMET	TerraQSAR
Data Quality	3	3	3	3	3
Chemical number	3	3	3	3	3
Descriptors/fragments	3	3	3	3	3
Algorithm	3	3	3	1	1
Applicability domain	3	3	3	3	3
Performance	3	3	3	3	3
Validation	3	3	3	3	3
Output	3	3	3	3	3
Cost	3	3	3	2	2
Total	27	27	27	24	24

For this endpoint there many models available. There are several reasons for this. This property is widely used both for ecotoxicological and toxicological properties, as an important component of other models. Indeed, logP is used to predict aquatic toxicity and bioconcentration factor, for the ecotoxicological and environmental behaviour properties, but it is also widely used for models of interest for human toxicology, like pharmacokinetics. Furthermore, there are data on more than 10,000 compounds. This prompted the development of many models, and many software houses have developed their own model, as a component of further models.

The performance of the models in general is quite good, also for the large number of data available. Many models give similar performance.

For this reason, a disciminant factor becomes the model transparency and cost. EPISuite resulted to be one of the most popular among stakeholders (ORCHESTRA project questionnaire). It offers advantages compared to other free tools: SPARC does not offer the batch mode, and VCCLAB refers to a quite old version of the software to calculate descriptors. Conversely, the US EPA model, EPISuite, is updated quite regularly.

Many commercial programs exist, such as ACD, TerraBase, ADMET Predictor, TOPKAT, Pallas. Marvin.

Thus, the model which seems preferable is EPISuite, even though similar results are expected using other programs. We will further evaluate some commercial programs, which nominally appear with very similar performance.

8.9.1 CARCINOGENICITY STUDY

Software (and developer):

Freely available:

- ✓ CAESAR project models (CAESAR consortium) http://www.caesar-project.eu/#;
- ✓ Lazar http://lazar.in-silico.de;
- ✓ OncoLogic (US EPA) http://www.epa.gov/oppt/sf/pubs/oncologic.htm;
- ✓ Toxtree http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/toxtree.

Commercial:

- ✓ DEREK (Lhasa Ltd) https://www.lhasalimited.org/index.php/Derek;
- ✓ MCASE/MC4PC http://www.multicase.com/products/prod01.htm;
- ✓ TOPKAT(Accelrys) http://accelrys.com/solutions/scientific-need/predictive-toxicology.html
- ✓ HazardExpert (CompuDrug)
- ✓ ADMET Predictor (Simulations Plus Inc.) http://www.simulations-plus.com/;
- ✓ Leadscope (Leadscope) http://www.leadscope.com/;
- ✓ MolCode Toolbox http://molcode.com/Hazard Expert (CompDrug)
- ✓

Overview of the evaluation of the QSAR models for carcinogenicity

	Caesar	Oncologic	Lazar	Toxtree	Topkat	Multicase	Derek	HE
Data Quality	3	3	3	3	3	3	3	3
Chemical number	2	2	2	2	2	2	2	2
Descriptors/fragments	3	3	3	3	3	3	3	3
Algorithm	3	3	3	3	1	1	2	1
Applicability domain	3	3	2	0	3	2	0	0
Performance	2	2	2	2	2	2	2	2
Validation	2	2	2	2	2	2	2	2
Output	2	2	2	2	2	2	2	2
Cost	3	3	3	3	2	1	1	2
Total	23	23	22	20	20	17	17	17

There are many criteria for which we assigned the same score, so they are not so critical. In this case the models typically refer to data of high quality. Thus we assigned the same score to all of them. In case of models based on fragment it is not fully appropriate to refer to "number" of chemicals in the training set, because there is not a training set, and the models are based on toxic fragments, derived from data on toxic/non toxci compounds. However, we assume that the fragments are derived on the basis of a large number of experimental data, and thus we assigned the same value as for the statistical models.

In case of commercial software the algorithm is not available, for commercial reasons, and only partial information is available. Derek provides more information and references on the data used to derive the evaluation. Thus we assigned a higher score to Derek, compared with other commercial software.

The applicability domain is typically a critical issue for models based on structural alerts, such as Derek, Toxtree and HazardExpert. Indeed, these models typically do not refer to a training set of compounds, and also the list of toxic fragments is not complete. Thus, it may be difficult to verify, in case of negative predictions, if the negative results is due to lack of knowledge on the toxic fragment. Lazar assigns as not reliable a quite large number of predictions, and Multicase provides many details which are subject to interpretation. For these reasons we reduced the score to these two programs.

The cost is the last criterion which shows different values.

Besides these 8 models, others exist. We can also evaluate the following three models, which seem promising: Leadscope, ADMET Predictor, Molcode.

8.4.1 IN-VITRO GENE MUTATION STUDY IN BACTERIA

Software:

Freely available:

- ✓ CAESAR project models (CAESAR consortium) http://www.caesar-project.eu/
- ✓ Lazar http://lazar.in-silico.de;
- ✓ T.E.S.T. http://www.epa.gov/nrmrl/std/cppb/qsar/, EPA;
- ✓ Toxtree http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/toxtree.

Commercial:

- ✓ ACD/Tox Suite (formerly ToxBoxes) http://www.acdlabs.com/products/pc_admet/tox/tox/;
- ✓ DEREK (Lhasa Ltd) https://www.lhasalimited.org/index.php/Derek;
- ✓ HazardExpert http://www.compudrug.com;
- ✓ MolCode Toolbox http://molcode.com/;
- ✓ TOPKAT(Accelrys) http://accelrys.com/solutions/scientific-need/predictive-toxicology.html

Overview of the evaluation of the QSAR models for mutagenicity

	Caesar	TEST	Lazar	Topkat	ACD	Molcode	Toxtree	Multicase	Derek	HE
Data Quality	3	3	3	3	3	3	3	3	3	3
Chemical number	3	3	3	3	3	3	2	3	2	2
Descriptors/fragments	3	3	3	3	3	3	3	3	3	3
Algorithm	3	3	3	1	1	1	3	1	2	1
Applicability domain	3	3	2	3	3	3	0	1	0	0
Performance	2	2	2	2	2	2	2	2	2	2
Validation	2	2	2	2	2	2	2	2	2	2
Output	2	2	2	2	2	2	2	2	2	2
Cost	3	3	3	2	2	2	3	1	1	2
Total	24	24	23	21	21	21	20	18	17	17

The evaluation is quite similar to what has been done previously for carcinogenicity.

The criteria with different scores will be considered here. The number of chemicals is quite large in case of statistical models for this endpoints, many thousands, while the number of structural fragments, and the associated chemicals, is lower than this figure. For this reason, Toxtree, Derek and HazardExpert got a lower score.

The transparency of the algorithm also provides different scores, because commercial models are not fully transparent.

For applicability domain, as in the case of models for carcinogenicity, the models based on structural alerts show limitations.

Furthermore, Lazar provides a not ideal tool for applicability domain, since in most of the cases the model recognizes as not reliable the prediction. Additionally, Lazar does not support the batch mode, which is a further criterion we identified within Deliverable 2, in case of more models to be evaluated.

Other promisig programs exist, like ADMET Predictor and Leadscope.

8.5. ACUTE TOXICITY

There are different acute toxicity in vivo models, depending on the exposure route. Oral exposure is the model more used, and for this reason there are more data available for this endpoint. As a consequence, most of the models have been developed for the oral exposure.

8.5.1. ACUTE TOXICITY - BY ORAL ROUTE

Software (and developer):

Freely available:

- ✓ T.E.S.T. http://www.epa.gov/nrmrl/std/cppb/qsar/, EPA

Commercial:

- ✓ ACD/Tox Suite (formerly ToxBoxes) http://www.acdlabs.com/products/pc_admet/tox/tox/;
- ✓ ADMET Predictor (Simulations Plus Inc.) http://www.simulations-plus.com/;
- ✓ MCASE/MC4PC http://www.multicase.com/products/prod01.htm;
- ✓ MolCode Toolbox http://molcode.com/;
- ✓ TerraQSAR™ - http://www.terrabase-inc.com/;
- ✓ TOPKAT(Accelrys) http://accelrys.com/solutions/scientific-need/predictive-toxicology.html

Overview of the evaluation of the QSAR models for LD50

LD50	T.E.S.T.	Topkat	Multicase
Data Quality	3	3	3
Chemical number	2	2	2
Descriptors/fragments	3	3	3
Algorithm	3	1	1
Applicability domain	3	3	2
Performance	2	2	2
Validation	2	2	2
Output	3	3	3
Cost	3	2	1
Total	24	21	19

The evaluation shows that all these models seem quite useful, even though this endpoint is not an easy one, and thus we cannot expect performance as in the case of physicochemical properties. Advantages for TEST are related to the fact that the tool is free and more transparent. Other programs seem also interesting: ADMET, ACD, TerraBASE, Molcode. Nominally, it is not easy to discriminate between programs which apparently promise similar performance.

9.1.1 SHORT-TERM TOXICITY TESTING ON INVERTEBRATES (DAPHNIA)

Software:

Freely available:

- ✓ Demetra http://www.demetra-tox.net/
- ✓ ECOSAR http://www.epa.gov/oppt/newchems/tools/21ecosar.htm
- ✓ T.E.S.T. http://www.epa.gov/nrmrl/std/cppb/qsar/index.html#TEST

Commercial:

- ✓ ADMET Predictor http://www.simulations-plus.com
- ✓ MolCode Toolbox http://molcode.com/;
- ✓ OASIS CATALOGIC http://www.oasis-lmc.org
- ✓ TerraQSAR™ - *Daphnia, Daphnia magna* 48-hr LC50 computation program http://www.terrabase-inc.com
- ✓ TOPKAT http://www.accelrys.com/products/topkat

Overview of the evaluation of the QSAR models for daphnia

	ECOSAR	T.E.S.T	Topkat
Data Quality	3	3	3
Chemical number	2	1	2
Descriptors/fragments	3	3	3
Algorithm	3	3	1
Applicability domain	3	3	3
Performance	3	2	3
Validation	2	2	2
Output	3	3	3
Cost	3	3	2
Total	25	23	22

In case of the models for daphnia, the number of chemicals is of a few hundreds. There are series of models, which may be candidate. The difference between the different models does not seem large. As in other cases, free models like ECOSAR offer the advantage of transparency and cost. The differences in the overall

evalution is not high, and other programs may be also evaluated, like TerraQSAR, ADMET Predictor and MolCode.

Demetra, which offer quite good performance, needs better evaluation regarding its applicability domain, because the model is based on pesticides.

9.1.3. SHORT-TERM TOXICITY TESTING ON FISH

Software (and developer):

Freely available:

- ✓ Demetra http://www.demetra-tox.net/
- ✓ ECOSAR http://www.epa.gov/oppt/newchems/tools/21ecosar.htm
- ✓ Lazar http://lazar.in-silico.de/
- ✓ T.E.S.T. http://www.epa.gov/nrmrl/std/cppb/qsar/index.html#TEST

Commercial:

- ✓ ADMET Predictor http://www.simulations-plus.com
- ✓ MolCode Toolbox http://molcode.com/
- ✓ OASIS http://www.oasis-lmc.org
- ✓ TerraQSAR™ - FHM http://www.terrabase-inc.com/
- ✓ TOPKAT http://www.accelrys.com/products/topkat

Overview of the evaluation of the QSAR models for fish

	ECOSAR	Lazar
Data Quality	3	3
Chemical number	2	2
Descriptors/fragments	3	3
Algorithm	3	3
Applicability domain	3	2
Performance	2	2
Validation	2	2
Output	3	3
Cost	3	3
Total	24	23

The situation for fish is quite similar to that for daphnia. The free models offer the advantage of the transparency and cost. Other candidate models may be TerraQSAR, MolCode, Topkat (commercial) and Demetra (free). The difference between the different models does not seem large.

9.2.DEGRADATION

9.2.1. BIOTIC DEGRADATION

9.2.1.1. READY BIODEGRADABILITY

Software:

Freely available:

- ✓ EPI Suite http://www.epa.gov/oppt/exposure/pubs/episuite.htm

Commercial:

- ✓ CATABOL, CATALOGIC http://oasis-lmc.org
- ✓ META , MultiCASE http://www.multicase.com/products/prod05.htm
- ✓ Molcode http://www.molcode.com/
- ✓ TOPKAT http://www.accelrys.com/products/topkat

degradation	EPISuite
Data Quality	3
Chemical number	2
Descriptors/fragments	3
Algorithm	3
Applicability domain	3
Performance	3
Validation	2
Output	3
Cost	3
Total	25

For degradation (property related to persistence) the models seem quite similar, and the discriminant appear once again due to the fact that preference may be given to more transparent models, also because the cost factor makes preferable the free EPISuite model.

Nevertheless, other candidate models will be possibly evaluated, like MolCode and Topkat.

9.3.2. BIOCONCENTRATION FACTOR (BCF)

Software:

Freely available:

- ✓ BCFWIN http://www.epa.gov/oppt/exposure/docs/episuitedl.htm
- ✓ CAESAR http://www.caesar-project.eu
- ✓ CQSAR http://www.biobyte.com/bb/prod/cqsarad.html
- ✓ DRAGON http://www.talete.mi.it/
- ✓ Fish model http://www.trentu.ca/academic/aminss/envmodel/models/Fish2.html
- ✓ Foodweb Model http://www.trentu.ca/envmodel/
- ✓ OECD QSAR Application toolbox http://www.oecd.org/env/existingchemicals/qsar
- ✓ T.E.S.T. http://www.epa.gov/nrmrl/std/cppb/qsar/index.html#TEST
- ✓ TAOBAC model http://www.trentu.ca/academic/aminss/envmodel/models/TAOv101.html

Commercial:

- ✓ ACD/LogDSuite http://www.acdlabs.com/products/pc_admet/physchem/physchemsuite/
- ✓ ASTER expert system http://www.epa.gov (not publicly available)
- ✓ MultiCASE http://multicase.com/
- ✓ OASIS CATABOL http://oasis-lmc.org/?section=software&swid=1

Scientific literature

- ✓ Toropov A.A., Toropova A.P., Lombardo A., Roncaglioni A., Benfenati E., Gini G. CORAL: Building up the model for bioconcentration factor and defining it's applicability domain. Eur J Med Chem, 46:1400-1403, 2011;
- ✓ Dimitrov, S.D., Mekenyan, O.G., Walker, J.D. Non-linear modeling of bioconcentration using partition coefficients for narcotic chemicals. SAR QSAR Environ Res, 13:(1):177-188, 2002;

Overview of the evaluation of the QSAR models for BCF

	Caesar	EPISuite	TEST	Meylan	Dimitrov	CORAL
Data Quality	3	3	3	3	2	3
Chemical number	2	2	2	2	2	2
Descriptors/fragments	3	3	3	3	3	3
Algorithm	3	3	3	3	3	2
Applicability domain	3	3	3	1	1	1
Performance	3	3	3	3	3	3
Validation	2	2	2	2	2	2
Output	3	3	3	3	3	3
Cost	3	3	3	3	3	3
Total	25	25	25	23	22	22

The models for bioconcentration factor (BCF) are quite numerous. This is due to the fact that hystorically a certain number of data was available, and that the property has an environmental relevance.

The results of the different models are not very different. There are several models which seem good candidate. As in other cases, free models show advantages. In this case there are also some relatively easy models which have been reported in the literature, and are based on one single parameter, logP. As we showed there are several programs suitable to calculate logP, and thus it may be easy to get the predicted value using these simple equations published in the literature to predict BCF.

A disadvantage is that somehow different results are expected, based on the different lopgP programs used. Another disadvantage is that these models do not offer a good clear basis to assess the applicability domain. Thus, programs like CAESAR, EPISuite and TEST seem preferable. Several commercial programs exist which may have similar performance.